

On the Heterogeneity of Digital Infrastructure in Entrepreneurial Ecosystems

Matthias Schulte-Althoff, Kai Schewina, and Daniel
Fürstenau
Department of Information Systems
Freie Universität Berlin
{matthias.schulte-althoff, kai.schewina,
daniel.fuerstenau}@fu-berlin.de

Gene Moo Lee
Accounting and Information Systems Division
UBC Sauder School of Business
gene.lee@sauder.ubc.ca

Abstract

Digital infrastructure represents for startups in entrepreneurial ecosystems an important asset but also a major risk. Drawing on studies about digital entrepreneurship and ecosystems, we examine the determinants of the heterogeneity of startups' tech stacks in ecosystems. Using publicly available data from the data aggregators Stackshare and Crunchbase, we identify popular endogenous categories in startups' tech stacks. Then we conduct a visual network analysis and a multivariate regression analysis, utilizing the identified technology categories to measure the heterogeneity of the startups' tech stacks. The analysis supports the propositions that firm age and increased funding are positively associated with tech stack heterogeneity, whereas funding rounds are negatively associated with tech stack heterogeneity. Implications of our findings on digital entrepreneurship and ecosystems are discussed.

1. Introduction

Studies of digital entrepreneurship have identified new needs and opportunities for entrepreneurs who rely heavily on digital technologies [1, 2]. Digital technologies make it easier to connect innovation efforts and enable more heterogeneous actors to work together [3]. In modern digital entrepreneurship ecosystems, the choice of technology holds both opportunities and risks for different ecosystem participants [4, 5].

The drivers of technology adoption decisions are interdependent and complex: Product characteristics, team skills, and knowledge are important factors. In addition, there are external factors related to the startup's ecosystem, such as trends [6], peers [7, 8], investors [9, 10], and sector effects [11]. The literature on technology adoption notes that technological interdependencies should be considered factors for

technology adoption [4, 6]. Adomavicius et al. [6] emphasize the spreading of technology along paths adopting a process theory perspective. However, there have been few studies of the quantification of determinants from the point of view of a macroscopic ecosystem. Nevertheless, this is an important aspect, as recent technology-related pitfalls show, such as the failure of the startup Hashtag Pirate after an API shutdown.¹ On the one hand, adopting homogeneous tech stacks can expedite the development of startups. But, on the other hand, heterogeneous tech stacks are important to build *robust* digital infrastructure. In this sense, this paper attempts to *identify the determinants of heterogeneity in tech stack adoption by startups, in the context of the entrepreneurial ecosystem in which they are embedded.*

Our study uses data sets collected from large-scale content aggregators Stackshare² and Crunchbase³. We conceptualize digital infrastructures as being organized in tech stacks, which we define as clusters of interrelated digital technologies. We consider the heterogeneity (or homogeneity) of the tech stack in relation to technology categories used by other startups in the ecosystem. Our main findings are that startups in an older age cohort and with more total funding tend to have more diverse tech stacks than younger and less-funded startups. In addition, startups that have collected funds over several rounds usually have a more homogeneous tech stack than those that have gone through fewer rounds of financing.

With the proposed methodology and empirical results, we contribute to the current research stream on collaboration and cooperation in entrepreneurial ecosystems with a focus on the use of digital infrastructure. In addition, our building block understanding of digital infrastructure as organized into tech

¹ <https://www.failory.com/interview/hashtag-pirate>

² <https://Stackshare.io/>

³ <https://www.Crunchbase.com/>

stacks is potentially valuable for companies becoming more distributed, embedded in dynamic ecosystems, and depending on co-creation of value while technologies become more layered and modular.

2. Digital Infrastructure of Startups in Entrepreneurial Ecosystems

In this section, we first define startups and entrepreneurial ecosystems. Then, we put forward our conceptualization of the digital infrastructure of startups as organized in tech stacks. Finally, we introduce technological heterogeneity of a startup's tech stack as the outcome variable of interest in our study, and we present potential determinants of tech stack heterogeneity driven by the entrepreneurial ecosystem in which a startup is embedded.

2.1 Entrepreneurial Ecosystems

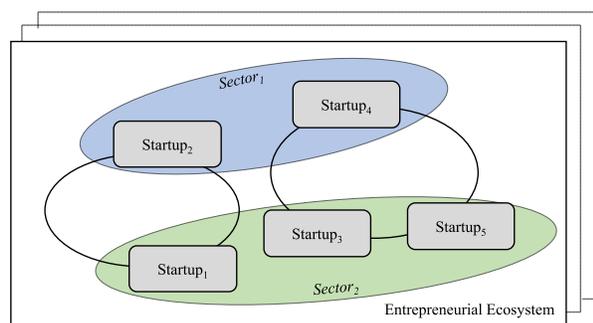


Figure 1. Selected elements of an entrepreneurial ecosystem

Entrepreneurs are individuals or companies that work on the discovery, evaluation, and exploitation of new business opportunities and are often regarded as change agents in their respective fields [12]. Digital entrepreneurs are seen as entities whose aim is to market, deliver, and support a digital product or service completely online [13]. In that way, digital entrepreneurs rely upon aspects of digital media and IT to pursue market opportunities [14]. Often they do so by making use of emerging digital infrastructure such as big data analytics, deep learning, virtual reality, IoT platforms, 3D printing, or cloud computing [15]. We apply the term ‘startup’ [12] to these IT-associated entrepreneurial companies.

Startups are embedded in entrepreneurial ecosystems comprising multiple types of actors often spanning across sector boundaries. Entrepreneurial ecosystems are a *set of interdependent organizations, including entrepreneurs, that pursue opportunities to create value-added products and services* [16, 17].

Current literature describes such collaboration as “open-system” orchestration [18]. As Guidici et al. [18] argue, opportunities reside not exclusively inside firms but can be co-created with other members of this ecosystem. While other types of actors besides startups (e.g., established companies, universities, government, and innovation intermediaries) also are important for such entrepreneurial ecosystems, this paper puts a focus on peer effects between startups as shown in Figure 1.

2.2 Understanding Digital Infrastructure as Organized in Tech Stacks

Not having endured the test of time and being less resilient, startups face many challenges. As Davidson and Vaast [14] put it, a digital startup has three interrelated types of opportunities: business, knowledge, and institutional. And as Steininger [12] asserts, their business model may be the main tool to seize opportunities, as it brings together economic and technological perspectives of innovations by adopting digital infrastructure as an infrastructural resource or as the product itself. As Davidson and Vaast [14] found, frameworks, digital artifacts, and digital platforms frequently serve as constituents of new venture ideas, while digital infrastructure serves as an enabler.

We focus on the intangible part of digital infrastructure as described by [3], which comprises software-based parts like stand-alone software tools and SaaS. The widespread use of such tech stacks has led to greater participation of actors in nearly all process phases of the ecosystem, including exploration, funding, and access to the market—a phenomenon named the “democratization” of entrepreneurship” [19].

Focusing on the tech stack, we can think of the elements as building blocks for the construction of larger components—the foundation of gradually enhancing the business model or the technological basis of the company [20]. This modular thinking is the core idea of microservices, which are independent processes that can communicate with each other [21]. This allows for executing single processes separately in contrast with many non-modular, monolithic structures. It has been argued that startup ventures can seize profit from this kind of split up value chain by incorporating well-established services into their company [22].

According to Nambisan’s literature review [1], technology has served as the context for empirical work in most research on digital entrepreneurship. Technology focuses either on opportunities for the

individual entrepreneurial venture (e.g., [21, 22]) or on the management, organization, and critical success factors for the entrepreneurial team (e.g., [23, 24]). In this work, we focus on technology as the core component from the point of view of an entrepreneurial ecosystem. In the next section, we briefly examine technology's role as a risk-bearer and provide one way of measurement.

2.3 Heterogeneity of Tech Stacks in Entrepreneurial Ecosystems

In this paper, we are interested in the *technological heterogeneity* of tech stacks in startups. We refer to technological homogeneity/heterogeneity as *the extent to which the tech stack used in a startup is concentrated (homogeneous) on few categories of technology versus diversified (heterogeneous) across multiple categories of technology compared to the tech stacks used in other startups*. Our definition thereby builds on a measure of entropy and considers different endogenously derived tech stack categories, which are aggregated into a single heterogeneity value.

There are many factors driving decisions for or against the use of a particular technology: team skills and knowledge, technology trends [6], geographical proximity, peer effects [7, 8], investor influence [9], [10] and sector effects [11] to name some important ones. In the field of technology adaptation, it was noted that technological interdependencies are also important for technology choice [4, 6], giving evidence that our chosen conceptualization is worth exploring. E.g., Adomavicius et al. [6] provide a tool for visualizing patterns of technological change over time, which are based on dominant technological roles.

The heterogeneity in digital infrastructure may be relevant to several actors in the entrepreneurial ecosystem. It was indicated that start-ups supported by early-stage and venture capital investors have less heterogeneous technology stacks than those with other types of investors. [27]. The authors argue that this may favor cluster risks of an investor's more homogeneous portfolio while yet simultaneously synergy effects may occur. Synergies may arise from the dissemination of information resulting from peer effects in a more closely aligned technological environment.

In the following, we introduce ecosystem-related factors that may influence the heterogeneity of tech stacks. So far, little attention has been paid to inter-organizational adaptation processes on the macro-level [28] and thus the alignment processes of digital infrastructure in ecosystems. However, prior research

documented the relationship between companies in an ecosystem and their corresponding digital infrastructure. For example, a study in the collaborative open-source software sector shows that the product of this community is much more modular than products of organizations with tighter boundaries [29].

A first factor we consider relates to the *number of funding rounds* a startup has received. Startups may accumulate funding over multiple rounds, maturing their operations, strategies, and business models. Those startups that make it to the later rounds (and, similarly, *seed stages*) may be more disciplined and thus may have better management and stricter controls. Recent studies show that venture capitalists play vital roles in the growth and the rate of success of new businesses [30]. Findings show that venture capitalists have an influence on the choices of both the CEO and other key personnel [28, 29]. They also show that venture capitalists provide value-adding services: Hellmann and Puri [33] provide empirical evidence that venture capitalists can support the building of human resources within the investees' organization. Hsu [34] points out substantial boosts in cooperative activity and performance associated with venture-capital-backed companies. Venture capitalists investing in the IT industry very often search for the next technological platform in which to invest [35]. Furthermore, there is a well-established research stream that shows accelerators and incubators form very tight relationships with their investees [33–37]. Thus, the literature points to a major role of early-stage investors in the development of new technologies. Overall, startups that have obtained multiple rounds of funding may be less heterogeneous in their tech stacks than those that do not obtain funding over many rounds. However, multiple funding rounds may also signal growth and maturity, so one could also argue that multiple rounds of funding could be associated with more diverse tech stacks.

Another factor we want to explore is the *total funding obtained* by a startup. The environment of startups is highly dynamic. Around 60% of startups do not survive the first five years, and the failure rate of startups funded by venture capitalists approximates 75% [41]. As early-stage startups often struggle with limited financial resources [42], investors often play a crucial role in their development. In particular, investors give startups the funding that enables them to scale products and the associated digital infrastructure. More funding may be a catalyst for diversifying the product portfolio and thus the tech stack used.

Table 1. Variable definition, operationalization, and distribution in our data set

Variable	Definition	Operationalization	Distribution
<i>Model variables</i>			
Technological heterogeneity	Software-sided technological heterogeneity regarding endogenous categories of tech stacks	Measure of entropy of the tech stack of a startup in respect to technology topics (metric)	Min: 0 (concentrated) Max: 1 (maximally diverse)
Funding round	The current funding round of a corresponding startup	Crunchbase assignment to startups (ordinal)	A (314) E (+) (52) B (239) unclassified (1,250) C (133) <i>No. of startups in</i> D (81) <i>parentheses</i>
Total funding	The sum of total funding given to a corresponding startup	Crunchbase assignment to startups (metric)	Median: 8,708,609 25% quartile: 1,000,000 75% quartile: 665,081,189
Founding date	The time that has passed since the founding of the company	Crunchbase assignment to startups; categorization in quarters (ordinal)	1: <i>young</i> (612), age < 5.5 years 2: <i>relatively young</i> (444), age > 5.5 years 3: <i>relatively old</i> (507), age > ~ 7 years 4: <i>old</i> (506), age > ~ 10 years <i>No of startups in parenthesis</i>
Common tech stack coverage	The extent to which startups use technologies together that other startups often use together.	Fulfilled association rules (base and add in basket) divided by the number of rules with present base technologies (metric)	Min: 0 (no rules fulfilled) Max: 1 (all possible rules fulfilled) Median: 0.452 25% quartile: 0.243 75% quartile: 0.655
<i>Additional descriptive variables</i>			
Startup	Startup engaging in the entrepreneurial ecosystem	Startups both in Stackshare and Crunchbase database (categorical):	<i>N</i> = 2,069 startups
Tech stack	Intangible part of the digital infrastructure (software and SaaS) used in startups	Stackshare assignment to startups (binary)	<i>N</i> = 1,283 technologies 0: used in startup 1: not used in startup
Sector	One or one of many business sectors a startup belongs to	Crunchbase assignment to a startup (categorical)	<i>N</i> = 45 sectors
Investor	Company investing in one or more of the startups	Presence of investment relationship between investor listed in Crunchbase and (minimum one) startup (binary)	<i>N</i> = 5449 investors • Venture capitalists (1695) • Early-stage investors (212) • Other (3542) <i>No of investors in parenthesis</i>

We also want to explore the role of the *founding date* of a startup and thus the age cohort to which a startup belongs. In particular, imprinting effects may play a role where startups in the same age cohort are similar in their level of heterogeneity [43]. This effect can be due to the availability of technologies and technological trends at the time. In terms of technology trend adoption, Gartner's Hype Cycle is used to show the typical process of innovation in terms of expectations and adoption [44]. Divided into five phases, the Hype Cycle indicates the typical adoption

process of innovative technologies—albeit from the perspective of digital technology and not from the perspective of an inter-entrepreneurial ecosystem. Over time, a bloating of tech stacks could occur where older age is associated with higher heterogeneity. Furthermore, this effect may be reinforced by complementarities such that the *coverage of common tech stacks* makes it increasingly likely that other related technologies are used in tandem. This, in turn, creates a path dependence, which makes it hard to break away from this base [42, 43]. Technologies

become increasingly intertwined and commonly used tech stacks may emerge. This may focus a startup's efforts on certain large tech stacks, thus leading to higher homogeneity.

3. Data and Methodology

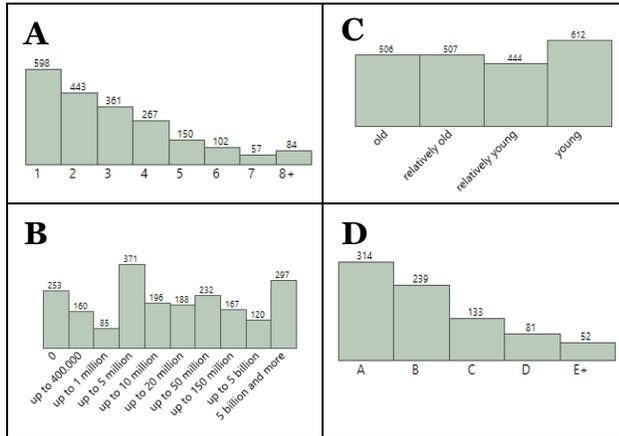


Figure 2. Distribution of funding round (A), funding total (B), founding date (C), and seed stage (D) in the startup data

This section describes data extraction, data reliability, and data analysis. The database used in the analysis was aggregated by means of the API of Crunchbase and collected from the Stackshare website.

Information on the startups and investments was acquired from Crunchbase, an open-source directory containing community-generated data on global technology companies and investors. The funding rounds provide information on the respective investor, invested firms, and amount invested. Technology datasets are extracted from Stackshare, an information aggregator that, like Crunchbase, is mainly curated by a community. More than 12,000 startups provide information about their tech stacks, which is not widely studied in the literature. To our knowledge, only Reeves et al. [47] operationalized technology data from Stackshare to analyze innovation as combining components to make new products.

Table 1 lists all considered variables as well as their operationalization and distribution from both data sources, and Figure 2 shows histograms of the distribution of the considered variables in our model. We considered the seed stage of a startup as another independent variable but, as Figures 2A and 2D

show, it is highly correlated to funding round (Pearson Correlation 0.68), thus we omitted it in our model.

Note that although Stackshare's notion of a startup is rather unclear, we can take into account Crunchbase's funding round information to control for this aspect. Finally, we are aware of the selection bias induced, as the sample only includes startups that are willing to enlist their tech stacks in Stackshare. Yet, since our focus lies on *digital* entrepreneurs, we believe the large sample from Crunchbase and Stackshare can provide a comprehensive picture of the entrepreneurial ecosystem.

Regarding the data reliability of Crunchbase and Stackshare, both data sources tightly monitor their data to correct inaccurate information. Furthermore, Crunchbase takes three means to ensure data curation [48]: First, the editors are part of the business. Second, Crunchbase uses machine-learning algorithms to compare data against publicly available information. Finally, data analysts recruited by Crunchbase take manual care of data validation. Being able to give basic trust to the data sources, we will subsequently present the methodology used to analyze the data in this paper.

The analysis phase consists of three main steps: First, a (visual) network analysis of tech stack co-occurrences utilizing association rule mining, second, an LDA topic modeling to categorize given tech stacks endogenously, and last, a regression analysis with technological heterogeneity operationalized by entropy values as the dependent variable. The aim of the regression is to investigate the influence of possible determinants of technological heterogeneity in entrepreneurial ecosystems.

4. Results

4.1 Visual Network Analysis

Basole et al. [49] study visual decision support for ecosystem analysis, and they find that network representation outperforms other frequently used methods such as matrices or lists for explorative visualization of complex systems. In this section, we consider two aspects in the context of network analysis: First, we examine the co-occurrences of technology by creating association rules and visualizing them in a network plot. Second, we reflect on the dependencies of funding and investors in the entrepreneurial ecosystem.

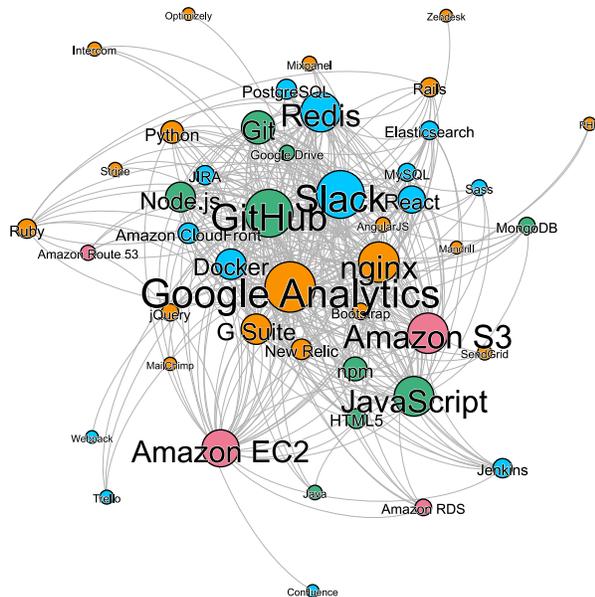


Figure 3. Network graphic of the co-occurrences of technologies based on association rule mining

For our first analysis, we took the technologies jointly used in startups' tech stacks as a basis for association rule mining—a common technique to find typical patterns of correlated goods. The goal is to find rules that describe which items are typically added to the current state of the shopping cart [50]. We conducted apriori rule mining with a minimum support of 0.15 and a minimum confidence of 0.5. This means that the rules must occur in at least 15% of the cases and that the rules must be fulfilled at least 50% of the cases [51]. The algorithm identified 1,623 rules covering 47 technologies. That means that 1,236 technologies are not included because of very low support for possible rules. A sensitivity analysis has shown that the amount of technologies does not significantly increase with a lower minimum support value.

The network plot in Figure 3 depicts the co-occurrence of technologies based on the found association rules. The node size depends on the degree, and the nodes have been visualized by four colors using a modularity clustering. The figure shows well-known digital technologies dominating the startup tech stacks. It is reasonable that big technology companies create technological dependencies and that most startups prefer common technological solutions.

Our second network analysis shows interdependencies between startups and investors, based on examples from the healthcare and finance sectors. Figure 4 presents these subsets of the entrepreneurial ecosystem depicting the finance sector

on the left and the health sector on the right, leaving out any startup with no connection to an investor.

The network of the entrepreneurial ecosystem depicts investors and startups as nodes. The edges connect investors and the corresponding startups in which they are investing. Technological heterogeneity between the nodes is represented as the edge weight using the Jaccard index, a well-known similarity index. The Jaccard similarity between a startup and an investor is calculated indirectly using the arithmetic mean of the similarity of the considered startup and the other startups supported by the examined investor. The resulting network is clustered by tech stack similarity and provides an explorative visualization.

We selected the financial and healthcare sectors as examples because they show particularly high modularity of the clusters, which indicates the use of similar tech stacks. The network indicates technological interdependencies in investor-startup relationships in terms of technological similarity that affect our independent variables *funding round* and *total sum of funding*.

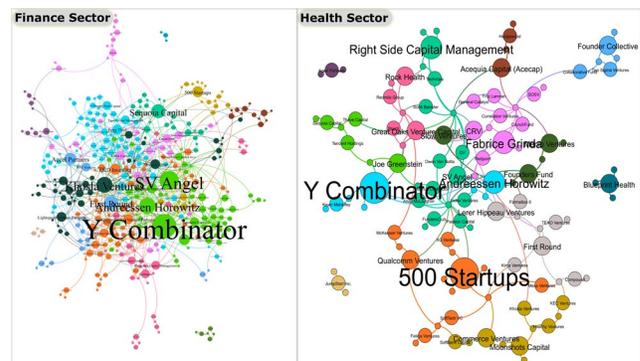


Figure 4. Network plot of the finance (left) and healthcare sectors (right), depicting investors and startups as nodes and investment relationships as edges. Technological homogeneity is used as edge weight.

4.2 Categories of Digital Infrastructure

To measure technological heterogeneity by means of an entropy index, we must categorize the tech stacks used by the startups. Our approach requires no manual classification; the categories are extracted endogenously from the Stackshare and Crunchbase data. We make use of Latent Dirichlet Allocation (LDA)—one of the most common topic modeling algorithms [52]. LDA is a text mining method that discovers “topics” from a large corpus of documents, building on the assumption that each word in each document is probabilistically drawn from the vocabulary of a topic discussed in that document.

Given a large collection of documents, the vocabularies of topics and the topics of the documents are jointly estimated.

Table 2. LDA results of the Stackshare technologies using Crunchbase company description data

Topic	Dim.	Top 5 Words
1	Web Dev.	rails, heroku, relic, sass, stripe
2	Web Dev.	php, wordpress, optimizely, server, adroll
3	Cloud Comp.	engine, webpack, cloud, es6, kubernetes
4	Cloud Comp.	elastic, elb, load, rds, route
5	App Dev.	kafka, java, scala, ansible, haproxy
6	App Dev.	rds, django, route, elastic, relic
7	Software / App Dev.	android, java, studio, intellij, sdk
8	Microsoft	microsoft, azure, visual, studio, net
9	Text Editing	sublime, text, php, drive, relic
10	Analytics	optimizely, mixpanel, relic, pingdom, labs

We carry out the algorithm on the Crunchbase dataset using *company descriptions*. Each firm has a document where the words are the tech stack names. The LDA model presents the common topics of tech stacks based on the co-occurrences of the stack names. Since the number of topics is a parameter that the user can choose, we tested the algorithm with different values (up to 100) and chose 10 as the value that best captures the technology topics. To illustrate that the results of the topic model comprehensively capture different dimensions of digital infrastructure, we list all 10 topics that LDA produces from our dataset in Table 2. Note that each topic is a distribution over all words in the vocabulary but that we show only the top five words for the sake of brevity.

4.3. Multivariate Analysis of the Heterogeneity of the Digital Infrastructure

To analyze the heterogeneity of tech stacks, we developed multivariate regression models. The dependent variable of the model's *Technological Heterogeneity* is the entropy measure that represents the diversification in the use of technology with respect to our endogenously generated technology topics in Section 4.2. An entropy value close to 1 therefore stands for a highly diverse tech stack, whereas an

entropy value close to 0 stands for a very homogeneous tech stack.

The models include the *logarithmized total funding*, *funding round*, and *founding date* as independent variables. In addition, the *common tech stack coverage*, i.e., the “association rule fulfillment quota,” was examined to check for robustness. This statistical measure returns the proportion of association rules we presented in Section 4.1 used in a startup’s tech stack.

Using OLS regression, we investigated the effect of each independent variable separately (Models 1 to 4) and then developed two multivariate models (Models 5 and 6). The lower *N* in Model 1 is the result of missing total funding values. We corrected for the missing values in the multivariate model. Table 3 provides the regression coefficients as well as the overall model diagnostics. All independent variables except for the *common tech stack coverage* show a highly significant impact on the outcome.

Whereas higher total funding leads to more technological heterogeneity, a greater number of funding rounds has the contrary effect. Considering the age of companies and taking the oldest startup companies as a baseline, the youngest ones (founding date group 0-1) were seen to have the highest technological heterogeneity. In comparison, the founding date group 1-2 is less heterogeneous. While the same is true for founding date group 2-3, interestingly, the oldest companies rise to approximately the same heterogeneity as group 1-2.

Models 5 and 6 each represent a multivariate model containing all independent variables—first without the common tech stack coverage variable and then with it. All values are close to their respective isolated impacts and significance. Because the metric *common tech stack coverage* has only a minimal impact on technological heterogeneity, we are likely to reject the hypothesis that the use of the most commonly co-occurring technologies influences the heterogeneity of the tech stack portfolio. We discuss the implications of our findings in the following section.

5. Discussion

We contribute to the literature on entrepreneurial ecosystems dealing with the evolution of infrastructure [2, 45] in an entrepreneurial environment. We are expanding this literature by taking into account digital infrastructure as the main object of consideration, incorporating a large public data set. Our findings tackle the question of what drives technological heterogeneity, considering determinants within these ecosystems.

Table 3. OLS regression analysis regarding the effects on *Technological Heterogeneity*

OLS Regression Results						
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Dependent Variable: Technological Heterogeneity						
<i>Intercept</i>	0.216***	0.510***	0.528***	0.441***	0.344***	0.375***
<i>Independent Variables</i>						
Total Funding [Log]	0.010*** (0.001)				0.011*** (0.001)	0.011*** (0.001)
Funding Round		-0.031*** (0.003)			-0.027*** (0.004)	-0.027*** (0.004)
Founding Date [1-2]			-0.132*** (0.023)		-0.073** (0.024)	-0.072** (0.024)
Founding Date [2-3]			-0.068** (0.024)		-0.057* (0.024)	-0.058* (0.024)
Founding Date [3-4]			0.061** (0.023)		0.084*** (0.023)	0.082*** (0.023)
Common tech stack coverage				-0.055* (0.030)		-0.068* (0.030)
N	1,816	2,069	2,069	2,069	2069	2069
R ²	0.033	0.034	0.042	0.002	0.087	0.089
Adjusted R ²	0.032	0.034	0.041	0.001	0.084	0.086
F-Statistic	61.14***	73.01***	30.51***	3.34	34.40***	29.58***
Standard errors in parentheses						
* significant with p < 0.1, ** p < 0.01, *** p < 0.001						

The first finding suggests that the more funding rounds the startups survive, the less heterogeneous their technological portfolio tends to be. It is well known that startups that survive many funding rounds are more robust and stable [53]. This stability may also extend to economic effects, e.g., the realization of economies of scale and being able to standardize.

Regarding funding, Gompers and Lerner [54] remind us that venture funding could be related to another unobserved factor of influence, the arrival of technological opportunities. In addition, the analysis shows that greater *total funding* leads to a more heterogeneous technological infrastructure. In this case, *total funding* may be a proxy for company size, since bigger companies use more technologies and may therefore be more diverse. The very small increase in technological homogeneity of the oldest group of companies may suggest that older companies have more difficulties standardizing technology [55].

Beyond our empirical findings, we make a methodical contribution to the understanding of digital infrastructure in entrepreneurship contexts [1]. We provide a data-driven method for operationalizing and endogenously categorizing the technology stacks of digital startups, based on a large amounts of publicly available data. Often, empirical analyses in this field rely on relatively small proprietary datasets or individual case studies, which limits their

generalization [12]. Through our study, the technology stacks of different digital startups can be compared, and similarities and differences can be identified on a broad basis.

Limitations of our study include the following point: The number of topics in LDA can be freely selected by the analyst. We manually tested various parameters for our analysis, which is enough to illustrate our general methodology. Nonetheless, an automatic approach as implemented by [56] could extend our method.

Further research can provide a concise sector-wise analysis within the ecosystem to check if startups from the same industries share the same technological footprint. Using an exploratory startup-investor network analysis with tech stack similarity as edges of the network, we have already been able to indicate that some sectors, such as finance and health care, are more prone to technological homogeneity. Similarly, the effect of other ecosystem actors can be subject to further investigation to answer the question whether investors influence the choices startups make about their digital infrastructure, as hinted by [27].

Furthermore, the findings of our study can be extended by using a longitudinal analysis to point out certain path dependencies or technology adoption patterns, e.g., for the field of AI: “That degree of influence translates into a strong dependence on

previous digital actions in the pattern of AI diffusion. European companies that have fully absorbed the previous set of digital technologies are 30 percent more likely to be first movers in AI adoption and use” [57].

6. Conclusion

We aimed to better understand the drivers of heterogeneity of digital infrastructure of startups in entrepreneurial ecosystems. Using a broad set of publicly available data incorporated by information aggregators on the web, we have shown from an ecosystem perspective that there are several measurable drivers of technological heterogeneity: Age, total funding, and number of funding rounds influence the heterogeneity of a startup’s digital infrastructure.

7. References

- [1] S. Nambisan, “Digital Entrepreneurship: Toward a Digital Technology Perspective of Entrepreneurship,” *Entrep. Theory Pract.*, vol. 41, no. 6, pp. 1029–1055, 2016.
- [2] S. Jarvenpaa and M. L. Markus, “Genetic Platforms and Their Commercialization: Three Tales of Digital Entrepreneurship,” in *51st Hawaii International Conference on System Sciences*, 2018.
- [3] K. Lyytinen, C. Sørensen, and D. Tilson, “Generativity in digital infrastructures,” *Routledge Companion to Manag. Inf. Syst.*, pp. 253–275, 2018.
- [4] Y. Yoo, O. Henfridsson, and K. Lyytinen, “The new organizing logic of digital innovation: An agenda for information systems research,” *Inf. Syst. Res.*, vol. 21, no. 4, pp. 724–735, 2010.
- [5] O. Henfridsson and Y. Yoo, “The Liminality of Trajectory Shifts in Institutional Entrepreneurship,” *Organ. Sci.*, vol. 25, no. 3, pp. 932–950, 2013.
- [6] G. Adomavicius, J. Bockstedt, A. Gupta, and R. Kauffman, “Making Sense of Technology Trends in the Information Technology Landscape: A Design Science Approach,” *Manag. Inf. Syst. Q.*, vol. 32, no. 4, Nov. 2008.
- [7] P. Almeida and B. Kogut, “Localization of knowledge and the mobility of engineers in regional networks,” *Manage. Sci.*, vol. 45, pp. 905–917, 1999.
- [8] G. L. Clark, M. Feldman, and M. Gertler, *The Oxford Handbook of Economic Geography Oxford*. Oxford University Press, 2000.
- [9] R. Grimaldi and A. Grandi, “Business Incubators and New Venture Creation: An Assessment of Incubating Models,” *Technovation*, vol. 25, pp. 111–121, 2005.
- [10] H. J. Sapienza and M. A. Korsgaard, “Procedural Justice in Entrepreneur-Investor Relations,” *Acad. Manag. J.*, vol. 39, no. 3, pp. 544–574, 1996.
- [11] P. Constantinides, O. Henfridsson, and G. G. Parker, “Introduction - Platforms and Infrastructures in the Digital Age - Semantic Scholar,” *Inf. Syst. Res.*, vol. 7047, pp. 1–20, 2018.
- [12] D. M. Steininger, “Linking information systems and entrepreneurship: A review and agenda for IT-associated and digital entrepreneurship research,” *Inf. Syst. J.*, 2018.
- [13] F. Zhao and A. Collier, “Digital Entrepreneurship: Research and Practice,” in *9th Annual Conference of the EuroMed Academy of Business*, 2016.
- [14] E. Davidson and E. Vaast, “Digital Entrepreneurship and its Sociomaterial Enactment,” in *43rd Hawaii International Conference on System Sciences*, 2010, pp. 1–10.
- [15] R. Beck, M. Avital, M. Rossi, and J. B. Thatcher, “Blockchain technology in business and information systems research,” *Bus. Inf. Syst. Eng.*, vol. 59, no. 6, pp. 381–384, 2017.
- [16] H. Van De Ven, “The development of an infrastructure for entrepreneurship,” *J. Bus. Ventur.*, vol. 8, no. 3, pp. 211–230, 1993.
- [17] B. Cohen, “Sustainable valley entrepreneurial ecosystems,” *Bus. Strateg. Environ.*, vol. 15, no. 1, pp. 1–14, Jan. 2006.
- [18] A. Giudici, “Open-System Orchestration As a Relational Source of Sensing Capabilities: Evidence From a Venture Association,” vol. 61, no. 4, pp. 1369–1402, 2018.
- [19] H. E. Aldrich, “The Democratization of Entrepreneurship? Hackers, Makerspaces, and Crowdfunding,” in *Academy of Management Annual Meeting, Philadelphia*, 2014.
- [20] W. B. Arthur, “The nature of technology: What it is and how it evolves,” *Free Press*, 2009.
- [21] N. Dragoni *et al.*, “Present and Ulterior Software Engineering,” M. M. und 38 M. Mazzara and B. Meyer, Eds. Springer International Publishing, 2017, pp. 195–216.
- [22] S. Tanev, “Is there a lean future for global startups?” *Technol. Innov. Manag. Rev.*, vol. 7, no. 5, pp. 6–15, 2017.
- [23] Y. Sarason, T. Dean, and J. F. Dillard, “Entrepreneurship as the nexus of individual and opportunity: A structuration view,” *J. Bus. Ventur.*, vol. 21, no. 3, pp. 286–305, 2006.
- [24] S. Shane and J. Eckhardt, “Handbook of entrepreneurship research: An interdisciplinary survey and introductions,” Z. J. Acs and D. B. Audretsch, Eds. Dordrecht, Kluwer Academic Publishers, 2003.
- [25] D. P. Forbes, P. S. Borchert, M. E. Zellmer-Bruhn, and H. J. Sapienza, “Entrepreneurial team formation: An exploration of new member addition,” *Entrep. Theory Pract.*, vol. 30, no. 2, pp. 225–248, 2006.
- [26] A. C. Klotz, K. M. Hmieleski, B. H. Bradley, and L. W. Busenitz, “New venture teams a review of the literature and roadmap for future research,” *J. Manage.*, vol. 40, no. 1, pp. 226–255, 2014.
- [27] M. Schulte-Althoff, K. Schewina, and D. Fürstenau, “A Risk Perspective on the Relation between

- Investors and the Digital Infrastructure of Startups,” in *Proceedings of the Seventy-ninth Annual Meeting of the Academy of Management*, 2019.
- [28] T. Oliveira and M. F. Martins, “Information technology adoption models at Firm Level: Review of literature,” *4th Eur. Conf. Inf. Manag. Eval. ECIME 2010*, vol. 14, pp. 312–322, 2010.
- [29] A. MacCormack, C. Baldwin, and J. Rusnak, “Exploring the duality between product and organizational architectures: A test of the ‘mirroring’ hypothesis,” *Res. Policy*, vol. 41, no. 8, pp. 1309–1324, 2012.
- [30] M. da Rin, T. Hellmann, and M. Puri, *A Survey of Venture Capital Research*, vol. 2, no. PA. Elsevier B.V., 2013.
- [31] T. Hellmann and M. Puri, “Venture capital and the professionalization of start-up firms: Empirical evidence,” *J. Finance*, vol. 57, no. 1, pp. 169–197, 2002.
- [32] O. Alexy, G. George, and A. J. Salter, “Cui bono? The selective revealing of knowledge and its implications for innovative activity,” *Acad. Manag. Rev.*, vol. 38, pp. 270–291, 2013.
- [33] T. Hellmann and M. Puri, “The interaction between product market and financing strategy: The role of venture capital,” *Rev. Financ. Stud.*, vol. 13, no. 4, pp. 959–984, 2000.
- [34] D. H. Hsu, “Venture Capitalists and Cooperative Start-up Commercialization Strategy,” *SSRN Electron. J.*, 2004.
- [35] D. Breznitz, C. Forman, and W. Wen, “The Role of Venture Capital in the Formation of a New Technological Ecosystem: Evidence from the Cloud,” *MISQ*, 2018.
- [36] T. Kohler, “Corporate accelerators: Building bridges between corporations and startups,” *Bus. Horiz.*, vol. 59, no. 3, pp. 347–357, 2016.
- [37] P. H. Phan, S. A. Mian, W. Lamine, B. Clarysse, M. Wright, and J. Van Hove, “A Look Inside Accelerators in the United Kingdom: Building Technology Businesses,” *Technol. Entrep. Bus. Incubation*, pp. 57–86, 2016.
- [38] A. Giudici, P. Reinmoeller, and D. Ravasi, “Open-System Orchestration as a Relational Source of Sensing Capabilities: Evidence from a Venture Association,” *Acad. Manag. J.*, vol. 61, no. 4, pp. 1369–1402, 2018.
- [39] S. Cohen, “What do accelerators do? Insights from incubators and angels,” *Innov. Technol. Governance, Glob.* 8(3-4), pp. 19–25, 2013.
- [40] D. L. Hoffman and N. Radojevich-Kelley, “Analysis of Accelerator Companies: An Exploratory Case Study of Their Programs, Processes, and Early Results,” *Small Bus. Inst. J.*, vol. 8, no. 2, 2012.
- [41] C. Nobel, “Why Companies Fail - and How Their Founders Can Bounce Back,” in *Harvard Business School*, 2011.
- [42] R. W. Smilor, “Commercializing technology through new business incubators,” *Res. Manage.*, vol. 30, no. 5, pp. 36–41, 1987.
- [43] A. L. Stinchcombe, “Social structure and organizations,” pp. 229–259.
- [44] T. Grundmeyer, “Adopting Technology: Using Student Qualitative Data and Gartner’s Hype Cycle,” *J. Educ. Train. Stud.*, vol. 2, no. 1, pp. 207–216, Nov. 2013.
- [45] W. Kremser and G. Schreyögg, “The Dynamics of Interrelated Routines: Introducing the Cluster Level,” *Organ. Sci.*, vol. 27, no. 3, pp. 698–721, Jun. 2016.
- [46] J. Sydow, G. Schreyögg, and J. Koch, “Organizational Path Dependence: Opening the Black Box,” *Acad. Manag. Rev.*, vol. 34, no. 4, pp. 689–709, Oct. 2009.
- [47] M. Reeves, T. Fink, R. Palma, and J. Harnoss, “Harnessing the Secret Structure of Innovation,” *MIT Sloan Manag. Rev.*, 2017.
- [48] R. Basole, H. Park, and R. O. Chao, “Visual Analysis of Venture Similarity in Entrepreneurial Ecosystems,” *IEEE Trans. Eng. Manag.*, vol. PP, pp. 1–15, 2018.
- [49] R. Basole, J. Huhtamäki, K. Still, and M. Russell, “Visual Decision Support for Business Ecosystem Analysis,” *Expert Syst. with Appl.* 65, vol. 65, 2016.
- [50] R. Agrawal and R. Srikant, “Fast Algorithms for Mining Association Rules,” *Proc. 20th Int. Conf. Very Large Data Bases VLDB*, vol. 1215, Aug. 2000.
- [51] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, “Efficient mining of association rules using closed itemset lattices,” *Inf. Syst.*, vol. 24, no. 1, pp. 25–46, Mar. 1999.
- [52] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet Allocation,” *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.
- [53] P. Gompers, A. Kovner, and J. Lerner, “Specialization and success: Evidence from venture capital,” *J. Econ. Manag. Strateg.*, vol. 18, no. 3, pp. 817–844, 2009.
- [54] P. Gompers and J. Lerner, “The Venture Capital Revolution,” *J. Econ. Perspect.*, vol. 15, no. 2, pp. 145–168, 2001.
- [55] J. D. Wareham, P. B. Fox, and J. L. Cano Giner, “Technology Ecosystem Governance,” *Organ. Sci.*, vol. 25, no. 4, pp. 1195–1215, 2014.
- [56] A. Lancichinetti, M. I. Sireer, J. X. Wang, D. Acuna, K. Körding, and L. A. N. Amaral, “A high-reproducibility and high-accuracy method for automated topic classification,” Feb. 2014.
- [57] J. Bughin, J. Seong, J. Manyika, L. Hämäläinen, E. Windhagen, and E. Hazan, “Notes from the AI frontier. Tackling Europe’s Gap in Digital and AI,” *McKinsey Glob. Inst.*, no. February, 2019.